


Requirements and Security

Security and the Grid Briefing

Rick St. Denis – Glasgow University

Rob Kennedy-FNAL

For the CDF DH group

Outline

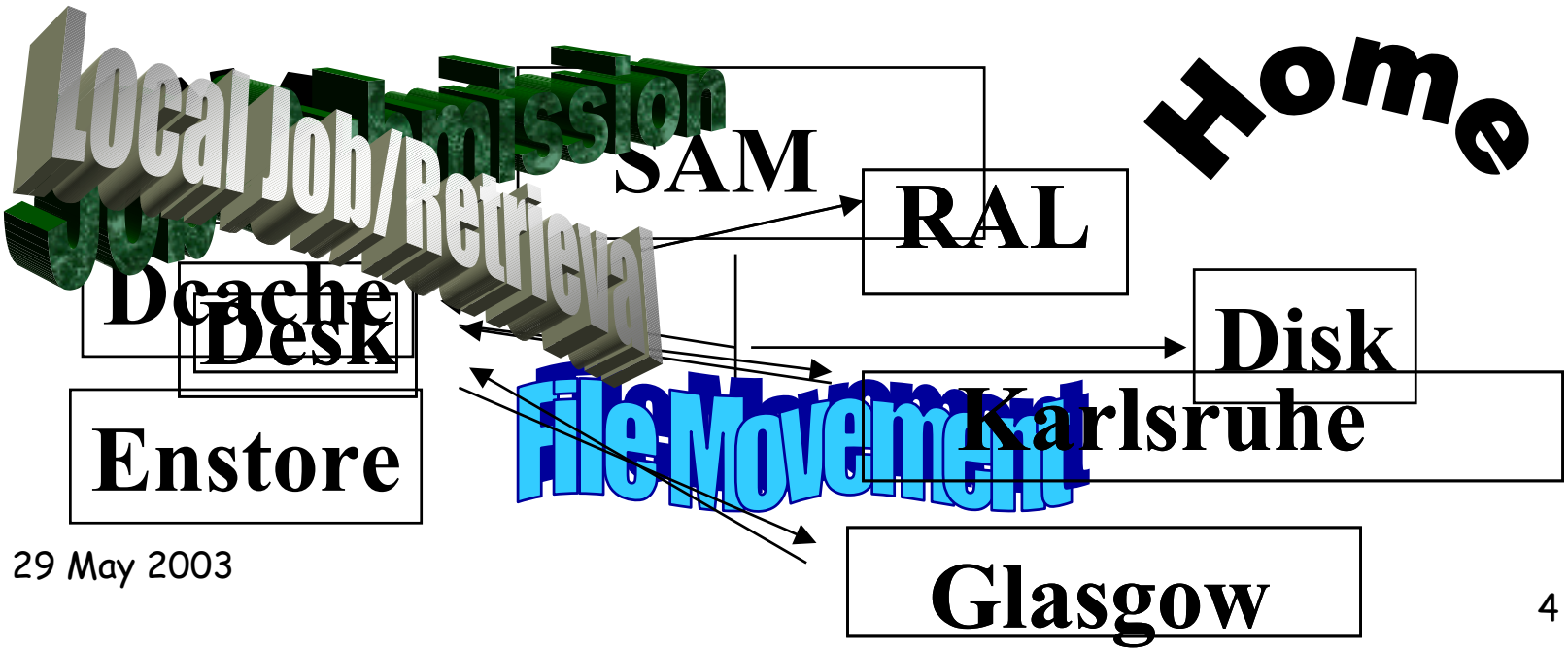
- DH Goals
- Implementation in CDF/D0
- Examples of use cases
- Expectations by end 2003
- Developments for the next 2 years
- Thoughts on Security and Conclusions

CDF and D0 DH Goals

- Offsite Computing
- MC Production
- Parallel Analysis
- User Metadata access
- Uniform access to data: tape and disk

The CDF DH Trident and D0 fork

- SAM: Data Catalog
- JIM: Global Batch Job Manager
- dCAF: Local Batch system



Examples Now

- 5 pc's with 20-30 Gig each make a 100 G cache. Can debug some datasets, accessing any (including raw data) up to keeping 100 files around. Desktops at FNAL; INFN
- Farm of PC's (10-20 at Oxford, Karlsruhe) with TB of disk nfs mounted, pc's use NAT to access FNAL, loading datasets and running from Dcache.


Ongoing Projects

- Short Range: What we expect by end 2003
  - Getting SAM as CDF's DH system
  - SAM/JIM deployment
- Long Range: Going into the next 2 years
  - Grid Operations
  - SAM/Grid Fusion
  - Security

Objectives of SAM-Grid

- JIM (Job and Information Management) complements SAM by adding job management and monitoring to data handling.
- Together, JIM + SAM = SAM-Grid
- Bring standard grid technologies (including Globus and Condor) to the Run II experiments.
- Enable globally distributed computing for DØ and CDF.

PPDG

GRIDPP UK Particle Physics

CDF II

the globus project™
www.globus.org

CDF Goals Now: SAM/dCaf/JIM

1. To run like the SC2002 Demo
2. Have the firewall problem overcome (dCAF problem)
3. To run physics analysis
4. To run production with storage to tape at fnal
5. To run stripping with storage to tape at fnal
6. To run Monte Carlo with storage to tape at fnal

JIM Deployment

Station	Uni	Items	Cpu	Disk	Admin	Share	Date
Oxford	Ox	1-6	20	7TB	Stonjek		5/31
Scotgrid2	Gla	1-6	128	7TB	BurgonLyon	20%	5/31
Glasgow	Gla	1-6	28	*	BurgonLyon	All	5/31
Sam	Fnal	1-6	60	2TB	St.Denis	Most	5/31
Trieste	Infn	6			Belforte	Italy	12/31
Toronto	Tor				Tafirout		5/31
Fzkka	Karls				Kerzel		5/31
Knu	Kor	1,6	12		Oh		
Ttu	Tt		8	3TB	Sill		5/31
Ucsd	Ucsd						

Short Range Milestones

JIM production at limited sites	6/1/03
MC Storage with AC++/SAM: demonstration	6/1/03
MC Production with AC++/SAM	10/1/03
JIM production at sites wanting JIM	10/1/03
CDF using SAM as basic DH system	10/1/03

Longer Term (next 2 years)

- Global Grid Operations
- Fusion With the Grid
- Security: design into the systems


Grid operations

- grow out of sam shifts at experiments
- controlled at tier 1 centers
- Stresses local manpower
  - ☞ Tier1 central monitoring according to subscription
    - Passive: provide tools for local monitoring
    - Active: subscription service. Central Lab support for HEP.
- Local implementation standardization and support
  - ➢ by major lab centers FNAL RAL CERN
    - ☞ - variety of packages for distribution with tailoring
    - ☞ - need to review the requirements satisfied by dcap, dccp, rootd, etc.
- Need to understand security implications


Fusion with Grid

- we know how to do this based on CDF/D0 join
- define database as an abstract concept: JTrumbo&Co
  - ☞ Allow for implementation of a single virtual RDBMS in
    - XML, Spitfire, Oracle, msql, mysql, postgres
    - Specify the requirements that can be met by each
- Again, need to understand security implications

What are the Security implications:

- peer to peer: complex and faulty!
- metadata query language, Bookkeeping on projects run, operations on datasets
- Virtual datasets and virtual dataset management from mc request of SAM

Security thoughts

- - essential to integrity of grid
  - drive by needs of user for security
  - illegal uses cited by RAL as a concern
  - model of
    - reasonable expectations from users
    - policing (PROactive not Reactive)
    - standardized destandardization: bbftp cookies in ups tailor at each site
    - allocation of systems with fast hack recovery: treat like a DB backup.
    - Orthogonal Security between sites has lost more time and system effort than a real hack attack! Any port (RAL) vs kerberos (FNAL).

Conclusions

- CDF and D0 are doing global computing and will increase the use of non-centralized computing
- Desktop and farm services to read data available now
- MC storage by October
- The user would like to get either
  - Results returned to the desktop
  - Results in a central repository where it can be retrieved.
- Security the essential consideration and causes more downtime than real hacking
- An expedient solution for the present requirements is needed
- dCaf gets the jobs running and back to you by being able to generate tickets: solutions in the grid are lacking
- Expansion to Grid and Tools Common with LHC will make the transition technically easier: solutions should endure